1. Project Name

   Prediction of Indiegogo Crowdfunding Success Rate

2. Description of Problem

   In today's startup world, whether it is a new product, or an existing service, or a personal goal to achieve which requires financial help, crowdfunding campaigns offer a way to raise debt and equity-free money. However, several articles pointed out a very low percentage of successful crowdfunding projects, and most articles compared different crowdfunding platforms to understand which one is better. In this particular problem, it is aimed to understand what projects are successful in a certain platform, Indiegogo, and how the success rate is affected by different parameters so that in the future, anyone who would like to raise funds for a purpose can decide whether to use Indiegogo, and how to use to conduct a successful campaign.

3. Project Clients

   This project's audience is considered to be anyone who would like to run a successful crowdfunding through Indiegogo. It is aimed that the audience will benefit from this study to understand projects in which categories are successful, how many updates or commenters are required to spread the word in other words how to interact with community, how the length of the description or the title affect the funds received, and whether geographic location where the project is launched on has any effect on the success rate.

4. Data to be used

   The dataset to be used is provided by the machine learning platform, bigml, through their website, and made it available by one of its users. The data consist of 15248 observations, a mixture of finished and unfinished crowdfunding projects. Each project has information regarding the number of updates on the project, the number of comments, the number of funders, success rate, goal, raised fund, category of the project, as well as geographical information, titles and description of the projects.

5. Outline of the Approach

   After the data are acquired, it is aimed to clean and explore the data. Once the data is ready for use, additional variables such as the number of the words in title of the project or the population of the geographical location will be created. The dataset will be divided into two as finished and unfinished projects. Finished projects will be divided into testing and training datasets. Different predictive modeling techniques will be conducted over training dataset. For example, either the prediction of success rate as a percentage shall be investigated, or whether the project will be successful (1) or not (0). Once the modeling is done, it will be tested on the testing dataset of already finished datasets. When the model with the highest accuracy (without overfitting) is selected, this will be sued to predict whether an unfinished project will be successful or not.

6. Deliverables

   A short report summarizing the description of the project and dataset, along with important findings and results will be delivered.